Plastid trnF pseudogenes are present in Jaltomata, the sister genus of Solanum (Solanaceae) : molecular evolution of tandemly repeated structural mutations

Poczai, Péter

2013

Poczai , P & Hyvönen , J 2013 , ' Plastid trnF pseudogenes are present in Jaltomata, the sister genus of Solanum (Solanaceae) : molecular evolution of tandemly repeated structural mutations ' , Gene , vol. 530 , no. 1 , pp. 143-150 . https://doi.org/10.1016/j.gene.2013.08.013

http://hdl.handle.net/10138/324958
https://doi.org/10.1016/j.gene.2013.08.013

Gene 530 (2013) 143–150

ACCEPTED AUTHORS MANUSCRIPT

Gene

Short Communication

Plastid *trn*F pseudogenes are present in *Jaltomata*, the sister genus of *Solanum* (Solanaceae): Molecular evolution of tandemly repeated structural mutations

Péter Poczai *, Jaakko Hyvönen

Plant Biology, Department of Biosciences, University of Helsinki, PO Box 65, FIN-00014 Helsinki, Finland

ARTICLE INFO

Article history:
Accepted 6 August 2013
Available online 17 August 2013

Keywords:
Chloroplast DNA
Gene duplications
Molecular recombination
Solanaceae
Tandem repeats

ABSTRACT

Extensive gene duplication arranged in a tandem array is rare in the plastome of embryophytes. Interestingly, we found pseudogene copies of the *trn*F gene in the genus *Jaltomata*, the sister genus of *Solanum* where such gene duplication has been previously reported. In each *Jaltomata* sequence available we found two pseudogene copies in close 5′-proximity to the original functional gene. The size of each pseudogene copy ranged between 17 and 48 bp and the anticodon domain was identified as the most conserved element. A common ATT$(G)_n$ motif is particularly interesting and its modifications were found to border the 3′ of the duplicated regions. Other motifs were partial residues, or entire parts of the T- and D-domains, and both domains proved to be variable in length among the pseudogenes identified. The residues of the 3′ and 5′ acceptor stem were not found among the copies. We further compared the newly discovered copies of *Jaltomata* with those ones previously described from *Solanum* and inferred phylogenetic relationships of the copies aligned. The evolution of *Solanum* copies, in contrast to *Jaltomata*, is hard to explain as resulting only in parsimonious changes since reticulate evolutionary patterns were detected among the copies. The dynamic evolutionary patterns of *Solanum* might be explained by possible inter- or intrachromosomal recombination.

1. Introduction

Pseudogenes are fragments of non-functional genomic DNA with high sequence similarity to normal functional genes (Chen et al., 2011; Nishishta et al., 2013). Since the plastome of embryophytes is known for its high degree of conservation in size, structure, gene content and linear order of genes (Palmer, 1991), such elements are rarely reported. Most structural mutations are composed of smaller indels (<10 bp) and the occurrence of large or complex rearrangements is sparse. Duplication of plastidic genes has been reported (Lindholm et al., 1991; Wakasugi et al., 1994) and transfer RNA (tRNA) genes might be involved in such processes via various molecular mechanisms (Hirasuka et al., 1989; Howe et al., 1988). However, the duplication of entire tRNA genes or their arrangement in a tandem array composed by multiple pseudogene copies is extremely rare in the plastome. Examples are known for the tandemly duplicated *trn*$T_{(GUU)}$ (Hirasuka et al., 1989; Howe et al., 1988), *trn*$S_{(GCU)}$ (Tsai and Strauss, 1989) and *trn*$Y_{(GUA)}$ (Hipkins et al., 1995). The duplication of the *trn*$F_{(GAA)}$ gene has also been described from Asteraceae (Vijverberg and Bachmann, 1999; Wittzell, 1999), Annonaceae (Pirie et al., 2007), Brassicaceae (Koch et al., 2007) and Juncaceae (Drábkova et al., 2004). The adjacent tRNA genes, *trn*$T_{(UGU)}$, *trn*$L_{(UAA)}$ and *trn*$F_{(GAA)}$ are located in the large single copy region of the chloroplast genome. They are co-transcribed and interspersed by two intergenic spacers and by a group I intron intercalated within the first and second exons of the *trn*$L_{(UAA)}$ gene. This region has been widely used in phylogenetic analyses of embryophytes and to address various questions of population genetics since the development of universal primers by Taberlet et al. (1991).

In our recent and ongoing molecular studies of the genus *Solanum* (Solanaceae), we encountered a tandem repeat comprising of two to four pseudogene copies upstream of the original *trn*F gene in four species of *Solanum* (Poczai and Hyvönen, 2011). We have characterized these structural duplications and shown that they consist of several highly structured motifs, which are partial residues, or entire parts of the anticodon, T- and D-domains of the original gene, but all lack the acceptor stems at the 5′- or 3′-end. We were further interested to evaluate the possible occurrence of complete or partial *trn*F pseudogenes in the genus *Jaltomata*. Recent molecular phylogenetic studies identified *Jaltomata* as the sister genus of *Solanum* (Hu and Saedler, 2007; Olmstead et al., 1999; Weese and Bohs, 2007) and the two genera make up together the tribe Solaneae (Olmstead et al., 2008). With about 60 species *Jaltomata* is a much smaller clade than *Solanum*, which

Abbreviations: EPT, equally parsimonious tree; IGS, intergenic spacer; MCP, multiple change point model; NNet, neighbor net; PSC, pseudogene copy; tRNA, transfer RNA; *trn*F, transfer RNA for phenylalanine; *trn*L, transfer RNA for leucine; *trn*T, transfer RNA for threonine.

* Corresponding author. Tel.: +358 9 191 57790; fax: +358 9 191 57788.
E-mail addresses: peter.poczai@gmail.com (P. Poczai), jaakko.hyvonen@helsinki.fi (J. Hyvönen).

0378-1119/$ – see front matter
http://dx.doi.org/10.1016/j.gene.2013.08.013

includes ~1400 species (Miller et al., 2011). The diversity of *Jaltomata* has only been recently discovered with various new species (e.g. Leiva et al., 2008; Mione et al., 2007) described. Many of these are narrow endemics and found in remote and historically poorly collected locations in the Andes of South America (Miller et al., 2011). For example at least six species live in the oases (Lomas) surrounded by hyperarid deserts scattered in the coastal deserts of northern Peru and Chile.

In the present study we report the discovery of *trn*F pseudogenes in sequences available in public genomic databases for *Jaltomata*. We describe the copies and their structural elements, and investigate the phylogenetic relationships of these copies. We compare nucleotide variation of pseudogenes in *Jaltomata* with its sister genus *Solanum*, and try to implicate the evolutionary mechanisms resulting in these tandem repeats.

2. Materials and methods

2.1. Sequence analyses and alignment

We carried out a search in NCBI nucleotide database (Benson et al., 2013) and downloaded available sequences for species of *Jaltomata* (Table 1). Putative pseudogene repeats were identified with screening using Repbase (Jurka, 2000) with the "mask pseudogenes" and "report simple repeats" options of the online tool CENSOR (Kohany et al., 2006). This was done to identify repetitive elements by comparing our sequences to known eukaryotic repeats and prototypic sequences stored in Repbase utilizing WU-BLAST. After repeats were identified in the *trn*L–F intergenic spacer (IGS) sequences, further structural *trn*F$_{(GAA)}$ gene elements or residues were annotated manually using the anticodon domain as reference. The annotated sequence alignment is shown in Fig. 1. Masked pseudogenic copies were further edited using Geneious v.4.8.5 (Biomatters Ltd.; Kearse et al., 2012). We used the *Nicotiana tabacum* L. complete chloroplast genome (Kunninimalayaan and Nielsen, 1997; NC_001879; bp positions 49,840 to 50,318) for comparisons and to determine the subunits of pseudogenic repeats as this species lacks such gene duplications. This sequence was used to select the corresponding loci of the *trn*L–*trn*F spacer region to annotate ambiguous sequence regions, and to ensure that our interpretations are based on homologous positions. Sequence break points were examined manually to determine the cut off points of pseudogenic copies and to identify bordering motifs. Isolated copies were aligned with MUSCLE (Edgar, 2004) as implemented in Geneious using default settings. The sequence alignment in FASTA format is available as Supplementary data 1.

2.2. BLAST search

In order to identify the parental regions of the pseudogene repeats we performed a nucleotide BLAST (Altschul et al., 1990) search (blastn) using the *Solanum tuberosum* L. plastome [accession number JF772170; Potato Genome Sequencing Consortium (2011)]. In this search PSC1–2 duplications were used as query sequences using the NCBI database.

Table 1
GenBank accession numbers of sequences used in this study.

Taxa	GenBank	Tribe	Copy number
Jaltomata Schltdl.			
J. auriculata (Miers) Mione	EU581006	Solaneae	2
J. grandiflora (B.L. Rob. Greenm.) D'Arcy, Mione & Tilton Davis	EU581007	Solanaceae	2
J. procumbens (Cav.) J.L. Gentry	AY098695	Solaneae	1[a,b]
J. procumbens (Cav.) J.L. Gentry	DQ180419	Solaneae	2
J. sinuosa (Miers) Mione	DQ180418	Solaneae	2
Solanum L.			
S. montanum L.	JN130369	Solaneae	2
S. physalifolium Rusby.	JN130371	Solaneae	4
S. sisymbriifolium Lam.	JN130366	Solaneae	5
S. trisectum Dunal	JN130370	Solaneae	2

[a] Partial pseudogenic copy at the 3′end.
[b] Missing original *trn*F gene.

2.3. Recombination detection

We tested for recombination using a Bayesian multiple change point model, DualBrothers (Minin et al., 2005; Suchard et al., 2003), implemented in the software package Geneious. Default settings were used with the following exceptions: chain length 220,000, subsampling frequency 100 and burn-in length 10,000. DualBrothers is a recombination detection algorithm based on a phylogenetic dual multiple change-point model (MCP; Suchard et al., 2003; Minin et al., 2005). The MCP model allows for changes in evolutionary relationships and rates across sites in a multiple sequence alignment by assuming that the sites separate into an unknown number of contiguous segments, each with possibly different topologies or mutation processes (McBride et al., 2009).

2.4. Phylogenetic analysis

We performed phylogenetic analyses with parsimony as an optimality criterion using the program Nona (Goloboff, 1994) within a WinClada (Nixon, 2002) shell. We performed five separate analyses (using processor time as a seed to randomize the order of the terminals) with the following settings: hold 3000 (holding defined number of trees), 100 replications (search performed with multiple tree-bisection–reconnection algorithm mult*max*), and hold/20 (keeping twenty starting trees for each replication). In addition we performed also a larger analysis by holding up to 30,000 trees (hold 30,000) but keeping only two starting trees for each replication (hold/2). Jackknife (Farris et al., 1996) support values were calculated using 1000 replications, with 10 search replications (multi*10) and with one starting tree per replication (hold/1).

2.5. Network analysis

A NeighborNet (NNet) analysis was carried out with SplitsTree v. 4.6 (Huson and Bryant, 2006). We conducted a NNet analysis in order to find out whether there are alternative explanations (i.e. recombination) for the observed similarities in different terminals. All calculations were based on uncorrected *p*-distances where splits (bipartitions) were represented as parallel edges. The resulting split graphs were filtered with a 95% threshold level to include only the splits with high support values.

3. Results

3.1. Structural characteristics of Jaltomata pseudogene copies

The length of the alignment containing the identified tandem repeats and part of the intergenic spacer region adjacent to the original *trn*F gene was 135 bp. We also aligned small part of the neighboring *trn*L–F IGS including the two putative promoter motifs in ultimate position plus some additional bases to see whether these regions have taken part in these duplications. BLAST search of the *S. tuberosum* plastome (accession number JF772170) using the *Jaltomata* copies as query sequences resulted in the *trn*F gene as the highest scoring element. The size of each pseudogenic copy (PSC) ranged between 17 and 48 bp and the anticodon domain was identified as the most conserved part. Anticodon regions of the *Jaltomata* PSC1 were nearly intact incorporating only two transitions at positions 101 and 109 (C → T), while PSC2 accumulated some transversion at aligned positions 88, 95 (G → T) and at 96 (A → C). We found a common ATT(G)$_n$ motif that is bordering the 3′ of the duplicated regions in all accessions similar as in Brassicaceae (Koch et al., 2005). This motif matches with a region found in close proximity to the original *trn*F gene in the tobacco genome (*N. tabacum*, NC_001879; 50,236–50,241 bp) and it can be found in all sequences.

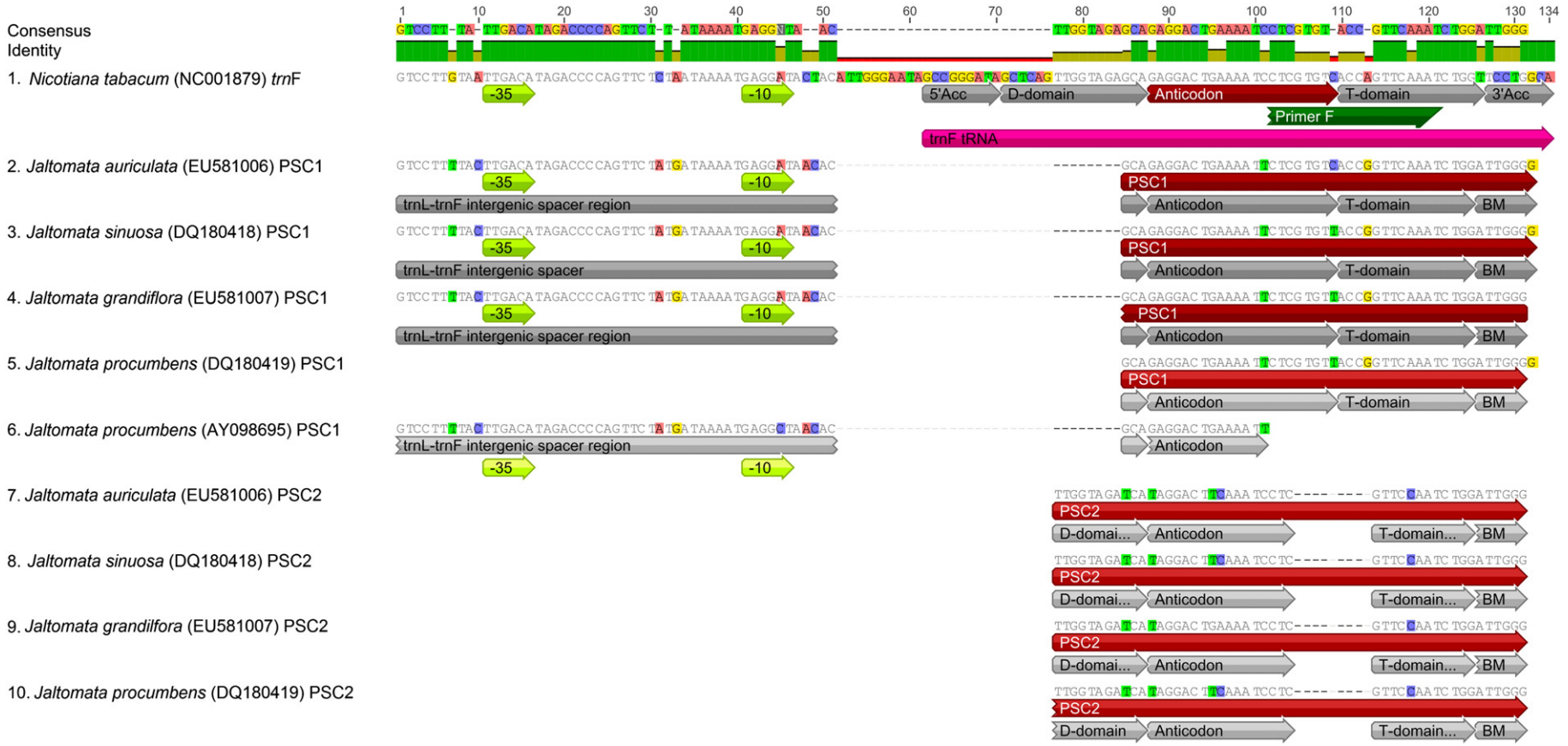

Fig. 1. Annotated sequence alignment of *Jaltomata* pseudogene copies and proximal bases. The sequence of *Nicotiana tabacum* is used as an outgroup comparison. Consensus identity scores are depicted on the top. Variable alignment positions are highlighted in color. Each pseudogene copy (PSC) is numbered subsequently and represented by *red* arrow. Various domains of the *trn*F gene namely the 5′acceptor stem, D-domain, anticodon, T-domain and 3′acceptor stem are depicted as the first entry in the alignment. Two promoter like motifs (−35 TTGACA/−10 GAGGAT) are annotated with *neon* arrows, while the *pink* arrow designates the original *trn*F gene. The primer *F* designed by Taberlet et al. (1991) is shown by the green arrow. Each copy shares a 3′ bordering motif (BM, *gray arrow*) ATT$(G)_{3-4}$ that is identical to a motif found in the intergenic spacer region of *N. tabacum* in close proximity to the *trn*F gene (alignment positions 52–57).

This was also proved to be an ideal cut off point between tandem copies. Other motifs were partial residues or entire parts of the T- and D-domains. The 3′ and 5′ acceptor stems were missing from all PSCs. The D-domain was almost completely missing from the first copies in all species and consisting of only a GCA residue that extends to more complete D-domain parts in the case of PSC2 accumulating a transversion at 85 (G → T). These copies also show a five bp deletion in the 3′ of the anticodon together with four missing bases from the 5′ of the T-domain as compared to PSC1. The T-domains incorporate only two mutations a transition (113, A → G) and a transversion (118, A → C). In *Jaltomata procumbens* (AY098695) only a minimum number of pseudogenes are present as this sequence is terminated with the pseudogene and not with the original *trn*F gene as revealed by our analysis. However, it is highly likely that the number of pseudogene copies would be two as observed in the other sequence (DQ180419). The incomplete sequence terminated with the first copy of a pseudogene is from the study of Garcia and Olmstead (2003) used for the phylogenetic analysis of tribe Anthocercideae (Solanaceae). However, this had no effect on their conclusions about the systematics of this group. Interestingly, PSC1 has an almost perfect primer binding site complementary with the *F* reverse primer (5′-ATT TGA ACT GGT GAC ACG AG-3′; complementary site ATT TGA ACT GGT GAC ACG AG; see Fig. 1 for details) described by Taberlet et al. (1991). This primer set ranging from *A* to *F* is commonly used to amplify the *trn*T–L spacer region. Due to this additional priming site we also observed a shorter faint band of ~418 bp besides the 550 bp main amplicon during our previous experiments with Solanaceae. If DNA is degraded the two can easily be mistaken for each other and this may have been the case in Garcia and Olmstead (2003).

3.2. Evolutionary relationships among Jaltomata and Solanum pseudogene copies

All parsimony analyses resulted in a set of equally parsimonious trees (EPTs) with the length of 88 steps. The ensemble consensus index (CI; Kluge and Farris, 1969) and ensemble retention index (RI; Farris, 1989) were both 0.55. The (strict) consensus trees of different analyses were identical, as illustrated in Fig. 2. The phylogenetic analyses suggested a monophyletic origin of the duplicated region in *Solanum* and *Jaltomata*. Four further synapomorphies separated the *Solanum* copies, although, different copy types were clearly indistinguishable except *Jaltomata* copy 2 that appears to be more closely related to the copies found in *Solanum*. The phylogeny revealed a high number of homoplastic characters mostly restricted to *Solanum*. This is also consistent

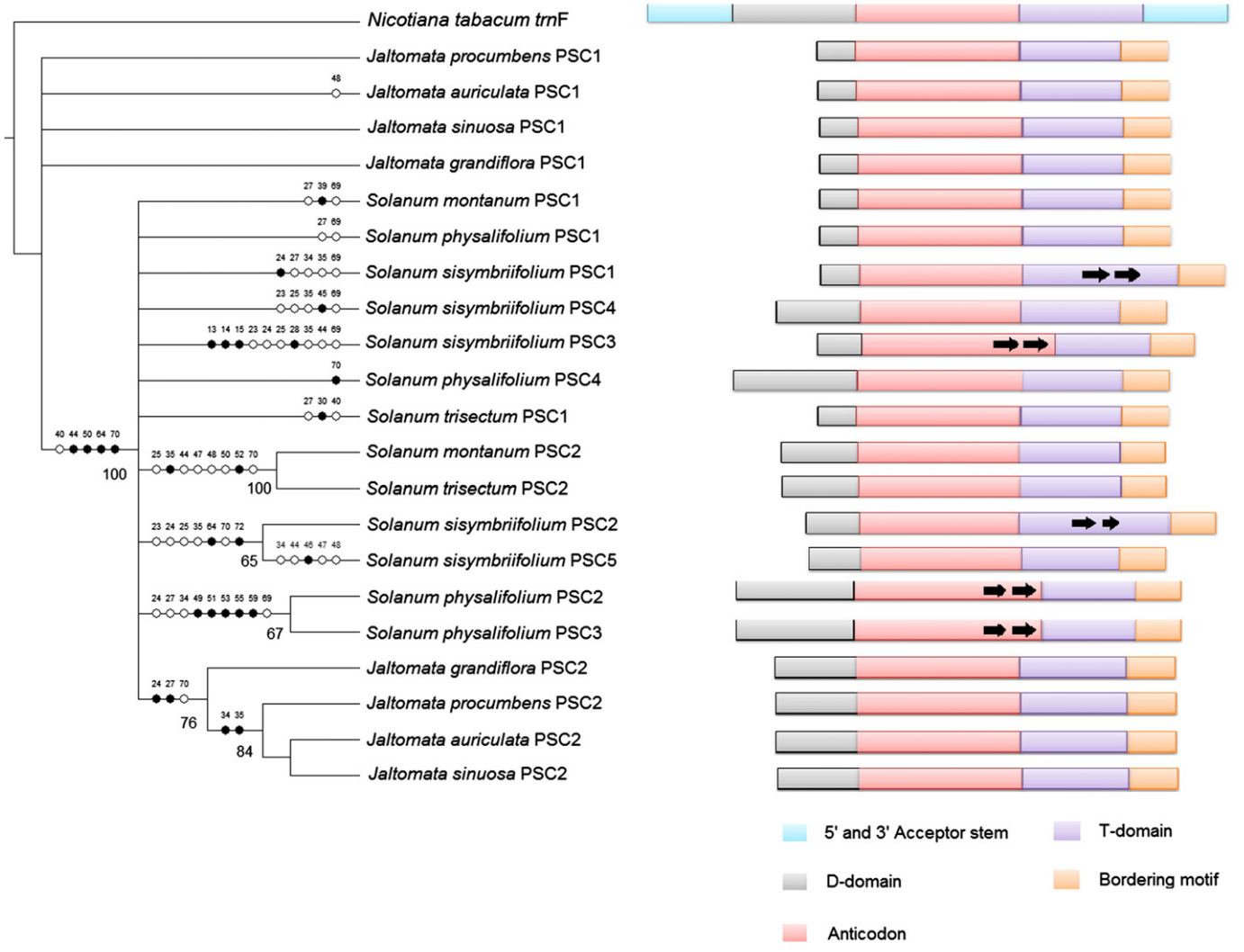

Fig. 2. Strict consensus of the EPT's obtained in the parsimony analyses of the *Jaltomata* and *Solanum* pseudogene copies. Synapomorphies are represented by black and homoplastic mutations as white dots. Numbers above the dots refer to the respective alignment position. Jackknife values >50 are shown under the branches. The trees have been rooted with the *trn*F gene of *Nicotiana tabacum*. On the right the schematic scheme of pseudogene copies is shown, where colored bars represent structural elements. Arrow signs depict sequence repeats described by Poczai and Hyvönen (2011).

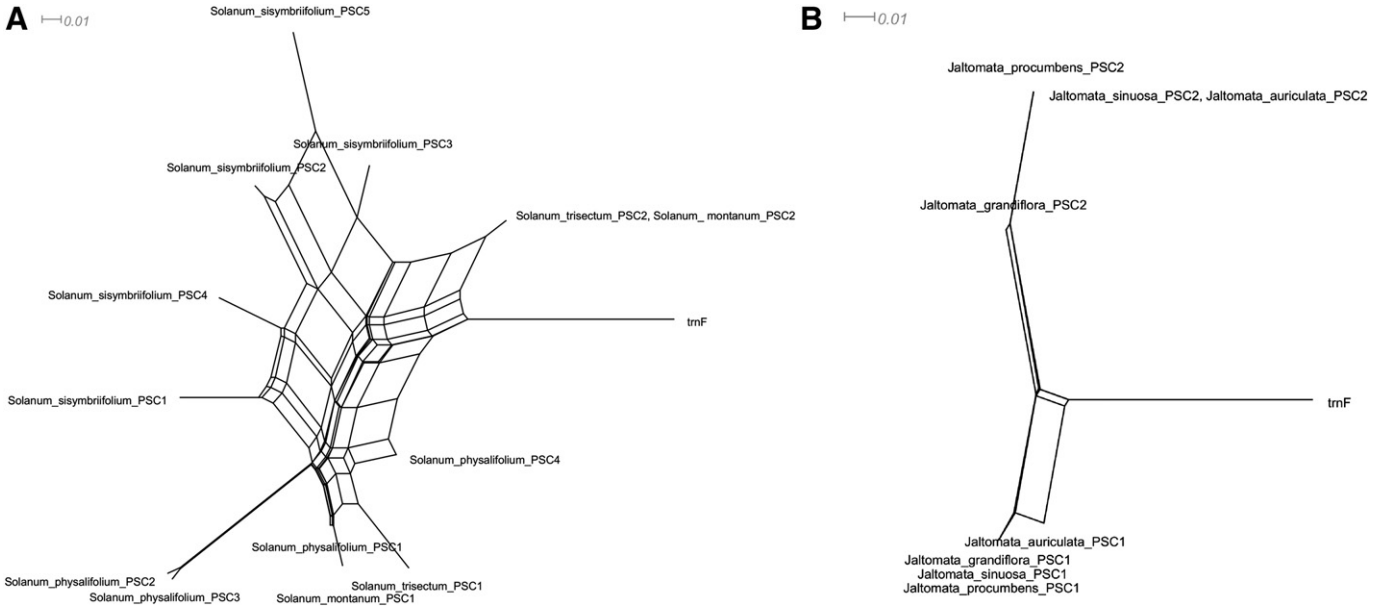

Fig. 3. Neighbornet (NNet) trees of pseudogene alignments inferred for *Solanum* (A) and *Jaltomata* (B) copies. The *trn*F sequence from *Nicotiana tabacum* is used as an outgroup.

with the constructed NNet tree which displays a complex evolutionary history for *Solanum* copies. In the NeighborNet analysis *Solanum* pseudogenes are highly divergent as compared to the outgroup *trn*F gene from tobacco (Fig. 3A). We detected many splits (bipartitions) between pseudogene copies indicating parallel substitutions, many being homoplasious in the parsimony analysis, which is consistent with patterns of intramolecular recombination and these were detected with DualBrothers. This method assumes that the evolutionary process is not independent and identically distributed at each site of the sequence and spatial phylogenetic variation along the sequences would be produced (Minin et al., 2005; Suchard et al., 2003). Instead, this model partitions the alignment into unknown number of segments. When appropriate sequences representing the parental one are included, the topology describing a putative recombination can vary and appear as changes in the most probable topology between the segments (Ferreira and Briones, 2012). Recombination breakpoints were found for the four analyzed *Solanum* species (Fig. 4). The breakpoint appeared near the end of PSC1 and 2 and start point for PSC3–5. We will not present an evolutionary scenario for *Solanum* pseudogene duplications as copy evolution in this genus seems to proceed with no clearly separated groups as can be seen also in the parsimony tree (Fig. 2) as well as in the NNet with many splits among groups (Fig. 3A). This differs from what can be observed in *Jaltomata* where copy 1 is linked with lateral lines to the outgroup *trn*F gene, and copies 1 and 2 also share some bipartitions from a 'splits rose' junction (Fig. 3B). This represents a scenario that copy 1 originated from the *trn*F gene and has later been transformed and duplicated resulting in copy 2. Statistical tests do not indicate any sign of recombination among *Jaltomata* pseudogene copies.

4. Discussion

The duplication of the *trn*F gene was first reported from *Microseris* D. Don and *Uropappus* Nutt. (Vijverberg and Bachmann, 1999) and *Taraxacum* Cass. (Wittzell, 1999) of Asteraceae. These tandem repeats are similar to those observed in *Jaltomata* and previously in *Solanum*. Almost the whole *trn*F gene is duplicated with high sequence similarity to the original plastid *trn*F gene. The expansion of the pseudogene tandem array in *Jaltomata* involves entire parts of the repeats without the disruptive breakdown in sequences. It might be that these nearly intact repeats were functional for some time in Solanaceae as hypothesized for Asteraceae. In *Microseris* and *Uropappus* 5′ and 3′ copies are so conserved that they might still be functional. The increased mutation rate detected in our case and the loss of acceptor stems at both ends may indicate the lack of functional constrains. The duplication of *Jaltomata*/*Solanum* copies also differ from reports published for Brassicaceae where only one major part (namely the anticodon domain) was involved in the initial duplication, thus it is hard to believe that such repeats are still functional (Dobeš et al., 2007).

4.1. The evolution of the pseudogene tandem repeat

The comparison of the *trn*L–F and sequences of *Jaltomata* and *Solanum* species with the plastome of *N. tabacum* suggested that the tandemly repeated *trn*F gene was not located at an inversion end point. We found no transposable element flanking the pseudogenes which could represent an alternative hypothesis for the origin of the tandem repeat. However, the length variation of the *trn*L–F intergenic spacer of different *Jaltomata* and *Solanum* species analyzed indicates that this region is prone to repeated changes in the *trn*F pseudogene copy number most likely starting from a duplicated sequence. Based on our analysis it is uncertain whether a duplicated or even a triplicated copy should be regarded as the ancient state as our taxonomic sampling is incomplete in both genera. Two to four tandemly repeated *trn*F copies were previously described from *Solanum* (Poczai and Hyvönen, 2011). We also found a fifth copy in *Solanum sisymbriifolium* previously overlooked correcting the number of copies for this species. In the sister genus *Jaltomata* all species available in GenBank for our investigation had two pseudogene copies. These copies shared mutations indicating that the duplication event appeared in the ancestor of *Jaltomata* species. Further mutation characteristic to some copies appeared after or during speciation events, rather than throughout intrachromosomal recombination between copies. We found no evidence that such mechanism would act with the copies present in *Jaltomata*. Additional shared nucleotide polymorphisms among *Jaltomata* and *Solanum* copies corroborate that at least some copies predate the separation of the two sister genera. It seems reasonable to expect that pseudogenes originate from a duplication event at the basal node uniting the two genera. This would imply that the duplication is rather ancient and may be present in other groups of Solanaceae like Capsiceae or Physaleae.

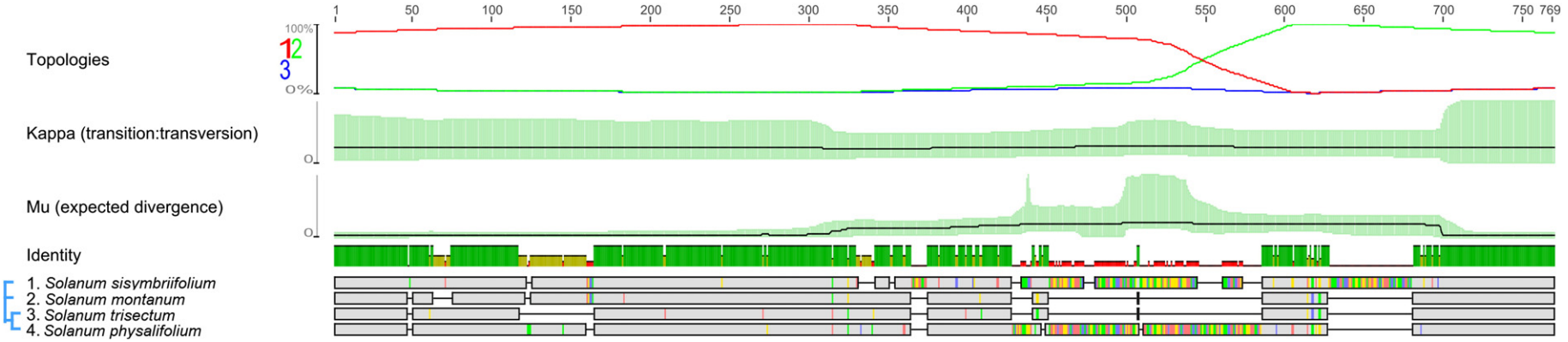

Fig. 4. Resulting recombination detection graphs from DualBrothers based on the multiple change-point (MCP) model among aligned nucleotide sequences of the complete *trn*L–F intergenic spacer region of the investigated *Solanum* species. Bases are represented by *gray* boxes while polymorphic sites are colored. Signs of possible recombination events are represented as changes in the topology graphs (first row) that appear at the most probable topology between the segments. Higher slopes in the μ graph are indicators of regions affected by recombination. The cross of the topological lines indicates recombination breakpoints.

4.2. Molecular mechanisms underlying the elongation of the tandem repeat

In *Solanum* pseudogene evolution seems to be highly dynamic compared to its sister genera *Jaltomata* with additional copies found in species that has been examined so far. Our analysis identified signs of recombination, which correspond to the short region of the *trn*L–F IGS region where the pseudogene duplications are exclusively intercalated. Previous studies have reported that structural mutations in the plastome accumulate whenever the first rearrangement has become incorporated (Palmer and Thompson, 1982) and previous rearrangements act as substrates for slippage mispairing and recombination (Cosner et al., 1997; Wolfson et al., 1991). The combined mechanisms of genetic exchange through intermolecular or intramolecular forms are likely to create such illegitimate recombination and highly reticulate evolutionary history as observed for *Solanum*. In *Arabidopsis lyrata* (L.) O'Kane & Al-Shehbaz such mechanisms were reported to be responsible for the dynamic variation observed in different haplotypes (Ansell et al., 2007). The length variation observed was later successfully used for phylogeographic analyses (Tedder et al., 2010). Ansell et al. (2007) also showed that copy number evolution is likely to follow iterative cycles of parallel expansion and contraction. If this is the case in *S. sisymbriifolium* and *Solanum physalifolium* sequencing effort with many specimens of each species would be required to infer the history of *trn*F duplications.

5. Conclusions and directions for further studies

Based on our results we assume that pseudogenes might be present also in other genera of Solanaceae. An extended search based on current sequence information available for Solanaceae should be carried out. Such search should be initiated by concentrating on *Solanum*, the largest genus that constitutes approximately half of the species (~1400) of the family. This huge genus seems to provide a good potential to explore the evolutionary dynamics of pseudogenes. Another important question arises from the fact that the exact phylogenetic origin of the *trn*F duplication is still uncertain. Further studies are also needed of the species such as *S. physalifolium* or *S. sisymbriifolium* where multiple copies of pseudogenes can be found. Larger populations of species like this should be studied in order to reveal whether parallel length changes are present. Such studies would elucidate whether unequal interchromosomal crossing-over is the driving force of expansion of the tandem repeats or if there are other mechanisms that underlie the evolution of such repeats.

Conflict of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

PP gratefully acknowledges support from the Marie Curie Fellowship Grant (PIEF-GA-2011-300186) under the seventh framework program of the European Union.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gene.2013.08.013.

References

Altschul, S., Gish, W., Miller, W., Myers, E., Lipman, D., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Ansell, S.W., Schneider, H., Pedersen, N., Grunmann, M., Russell, S.J., Vogel, J.C., 2007. Recombination diversifies chloroplast *trn*F pseudogenes *Arabidopsis lyrata*. J. Evol. Biol. 20, 2400–2411.

Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Sayers, E.W., 2013. GenBank. Nucleic Acids Res. 41, D36–D42.

Chen, S.-M., Ma, K.-Y., Zeng, J., 2011. Pseudogene: lessons from PCR bias, identification and resurrection. Mol. Biol. Rep. 38, 3709–3715.

Cosner, M.E., Jansen, R.K., Palmer, J.D., Downie, S.R., 1997. The highly rearranged chloroplast genome of *Trachelium caeruleum* (Campanulaceae): multiple inversions, inverted repeat expansion and contraction, transposition, insertions/deletions, and several repeat families. Curr. Genet. 31, 419–429.

Dobeš, C., Keifer, C., Keifer, M., Koch, M.A., 2007. Plastidic *trn*F_{UUC} pseudogenes in North American genus *Boechera* (Brassicaceae): mechanistic aspects of evolution. Plant Biol. 9, 502–515.

Drábkova, L., Kirschner, J., Vlček, Č., Paček, V., 2004. *Trn*L–*trn*F intergenic spacer and *trn*L intron define major clades within *Luzula* and *Juncus* (Juncaceae): importance of structural mutations. J. Mol. Evol. 59, 1–10.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.

Farris, J.S., 1989. The retention index and rescaled consistency index. Cladistics 5, 417–419.

Farris, J.S., Albert, V., Källersjö, M., Lipscomb, D., Kluge, A.G., 1996. Parsimony jackknifing outperforms neighbor-joining. Cladistics 12, 99–124.

Ferreira, R.C., Briones, M.R.S., 2012. Phylogenetic evidence based on *Trypansoma cruzi* nuclear gene sequences and information entropy suggest that inter-strain intragenic recombination is a basic mechanism underlying the allele diversity of hybrid strains. Infect. Genet. Evol. 12, 1064–1071.

Garcia, V.F., Olmstead, R.G., 2003. Phylogenetics of tribe Anthocercideae (Solanaceae) based on *ndh*F and *trn*L/F sequence data. Syst. Bot. 28, 609–615.

Goloboff, P.A., 1994. NONA: a tree searching program. Program and documentation. www.cladistics.com/aboutNona.htm.

Hipkins, V.D., Marshall, K.A., Neale, D.B., Rottmann, W.H., Strauss, S.H., 1995. A mutation hotspot in the chloroplast genome of a conifer (Douglas-fir: *Pseudotsuga*) is caused by variability in the number of direct repeats from partially duplicated tRNA gene. Curr. Genet. 27, 527–579.

Hirasuka, J., et al., 1989. The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. Mol. Gen. Genet. 217, 185–194.

Howe, C.J., Barker, R.F., Bowman, C.M., Dyer, T.A., 1988. Common features of three inversions in wheat chloroplast DNA. Curr. Genet. 13, 343–349.

Hu, J.Y., Saedler, H., 2007. Evolution of the inflaxed calyx syndrome in Solanaceae. Mol. Biol. Evol. 24, 2443–2453.

Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254–267.

Jurka, J., 2000. Repbase update: a database and an electronic journal of repetitive elements. Trends Genet. 9, 418–420.

Kearse, M., et al., 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649.

Kluge, A.G., Farris, J.S., 1969. Quantitative phyletics and the evolution of anurans. Syst. Zool. 18, 1–32.

Koch, M.A., et al., 2005. Evolution of the *trn*F(GAA) gene in *Arabidopsis* relatives and the Brassicaceae family: monophyletic origin and subsequent diversification of a plastidic pseudogene. Mol. Biol. Evol. 22, 1032–1043.

Koch, M.A., Dobeš, C., Kiefer, C., Schmickl, R., Klimeš, L., Lysak, M.A., 2007. Supernetwork identifies multiple events of plastid *trn*$F_{(GAA)}$ pseudogene evolution in the Brassicaceae. Mol. Biol. Evol. 24, 63–73.

Kohany, O., Gentles, A.J., Hankus, L., Jurka, J., 2006. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. BMC Bioinf. 7, 474.

Kunninimalayaan, M., Nielsen, B.L., 1997. Fine mapping of replication origins (ori A and ori B) in *Nicotiana tabacum* chloroplast DNA. Nucleic Acids Res. 25, 3681–3686.

Leiva, G., Mione, S.T., Yacher, L., 2008. Dos nuevas especies de *Jaltomata* Schlechtendal (Solanaceae) del Norte del Perú. Arnoldia 15, 185–196.

Lindholm, J., Szmidt, A., Gustafsson, P., 1991. Duplication of the *psb*A gene in the chloroplast genome of two *Pinus* species. Mol. Gen. Genet. 226, 345–352.

McBride, A.J.A., et al., 2009. Genetic diversity of Leptospiral immunoglobulin-like (Lig) genes in pathogenic *Leptospira* spp. Infect. Genet. Evol. 9, 196–205.

Miller, R.J., Mione, T., Phan, H.-L., Olmstead, R.G., 2011. Color by numbers: nuclear gene phylogeny of *Jaltomata* (Solanaceae), sister genus to *Solanum*, supports three clades differing in fruit color. Syst. Bot. 36, 153–162.

Minin, V.N., Fang, F., Dorman, K.S., Suchard, M.A., 2005. Dual multiple change-point model leads to more accurate recombination detection. Bioformatics 21, 3034–3042.

Mione, T., Leiva, S.G., Yacher, L., 2007. Five new species of *Jaltomata* (Solanaceae) from Cajamarca, Peru. Novon 17, 49–58.

Nishishta, K., Sakai, E., Okamoto, K., Tsukuba, T., 2013. Structural and phylogenetic comparison of napsin genes: the duplication, loss of function and human-specific pseudogenization of napsin B. Gene 517, 147–157.

Nixon, K.C., 2002. Winclada. Version 1.00.08. www.cladistics.com/aboutWinc.htm Ithaca, NY, USA.

Olmstead, R.G., Sweere, J.A., Spangler, R.E., Bohs, L., Palmer, J., 1999. Phylogeny and Provisional Classification of the Solanaceae Based on Chloroplast DNA. In: Nee, M., Symon, D., Lester, R.N., Jessop, J. (Eds.), Kew, Royal Botanic Gardens, UK, pp. 111–137.

Olmstead, R.G., Bohs, L., Migid, H.A., Santiago-Valentin, E., Garcia, V.F., Collier, S.M., 2008. A molecular phylogeny of the Solanaceae. Taxon 57, 1159–1181.

Palmer, J.D., 1991. Plastid chromosomes: structure and evolution. In: Bogorad, L., Vasil, I.K. (Eds.), Cell Culture and Somatic Cell Genetics in Plants. The Molecular Biology of Plastids, 7. Academic Press, San Diego, pp. 5–53.

Palmer, J.D., Thompson, W.F., 1982. Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. Cell 29, 537–550.
Pirie, M.D., Vargas, M.P.B., Botermans, M., Bakker, F.T., Chatrou, L.W., 2007. Ancient paralogy in the cpDNA *trn*L–F region in Annonaceae: implications for plant molecular systematics. Am. J. Bot. 94, 1003–1016.
Poczai, P., Hyvönen, J., 2011. Identification and characterization of plastid *trn*$F_{(GAA)}$ pseudogenes in four species of *Solanum* (Solanaceae). Biotechnol. Lett. 33, 2317–2323.
Potato Genome Sequencing Consortium, 2011. Genome sequence and analysis of the tuber crop potato. Nature 475, 189–195.
Suchard, M.A., Weiss, R.E., Dorman, K.S., Sinsheimer, J.S., 2003. Inferring spatial phylogenetic variation along nucleotide sequences: a multiple change-point model. J. Am. Stat. Assoc. 98, 427–437.
Taberlet, P., Gielly, L., Pautou, G., Bouvet, J., 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol. Biol. 17, 1105–1109.
Tedder, A., Hoebe, P.N., Ansell, S.W., Mable, B.K., 2010. Using chloroplast trnF pseudogenes for phylogeography in *Arabidopsis lyrata*. Diversity 2, 653–678.
Tsai, C.-H., Strauss, S.H., 1989. Dispersed repetitive sequences in the chloroplast genome of Douglas-fir. Curr. Genet. 16, 211–218.
Vijverberg, K., Bachmann, K., 1999. Molecular evolution of tandemly repeated *trn*F(GAA) gene in the chloroplast genomes of *Microseris* (Asteraceae) and the use of structural mutations in phylogenetic analysis. Mol. Biol. Evol. 16, 1329–1340.
Wakasugi, T., Tsudzuki, J., Ito, S., Nakashima, K., Tsudzuki, T., Sugiura, M., 1994. Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. Proc. Natl. Acad. Sci. USA 91, 9794–9798.
Weese, T.L., Bohs, L., 2007. A three-gene phylogeny of the genus *Solanum* (Solanaceae). Syst. Bot. 32, 445–463.
Wittzell, H., 1999. Chloroplast DNA variation and reticulate evolution in sexual and apomictic sections of dandelions. Mol. Ecol. 8, 2023–2035.
Wolfson, R., Higgins, K.G., Sears, B.B., 1991. Evidence for replications slippage in the evolution of *Oenothera* chloroplast DNA. Mol. Biol. Evol. 8, 709–720.